

6. The Biological Justification of Ethics: A Best-Case Scenario

Social and behavioral scientists - that is, students of human nature - nowadays hardly ever use the term 'human nature'. This reticence reflects both a becoming modesty about the aims of their disciplines and a healthy skepticism about whether there is any one thing really worthy of the label 'human nature'.

For some feature of humankind to be identified as accounting for our 'nature', it would have to reflect some property both distinctive of our species and systematically influential enough to explain some very important aspect of our behavior. Compare: molecular structure gives the essence or the nature of water just because it explains most of its salient properties. Few students of the human sciences currently hold that there is just one or a small number of such features that can explain our actions and/or our institutions. And even among those who do, there is reluctance to label their theories as claims about 'human nature'.

Among anthropologists and sociologists, the label seems too universal and indiscriminant to be useful. The idea that there is a single underlying character that might explain similarities threatens the differences among people and cultures that these social scientists seek to uncover. Even economists, who have explicitly attempted to parlay rational choice theory into an account of all human behavior, do not claim that the maximization of transitive preferences is 'human nature'.

I think part of the reason that social scientists are reluctant to use 'human nature' is that the term has traditionally labeled a theory with normative implications as well as descriptive ones. Any one who propounds a theory of human nature seems committed to drawing conclusions from what the theory says is the case to what *ought* to be the case. But this is just what twentieth-century social scientists are reluctant to do. Once the lessons of David Hume and G.E. Moore were well and truly learned among social scientists, they surrendered the project (associated with the 'moral sciences' since Hobbes) of deriving 'ought' from 'is'.¹

¹ David Hume, *A Treatise of Human Nature*, ed. L.A. Selby-Bigge (Oxford: Clarendon Press, 1888), bk. II; G.E. Moore, *Principia Ethica* (Routledge and Kegan Paul, 1907).

Do not quote without approval of author alexrose@duke.edu

The few scientists who have employed the term ‘human nature’ do draw evaluative conclusions from their empirical theories. The best recent examples of such writers are sociobiologists like E.O. Wilson, eager to extend the writ of evolutionary biology to include both the empirical study of humans and the foundations of their moral philosophy.²

It is relatively easy to offer a review and philosophical critique of the excesses that are bound to creep into evolutionary biologists’ attempts to transcend the traditional limits of their discipline. But more useful than still another catalog of sociobiological foibles would be a sympathetic examination of the best we might hope for from the application of evolutionary biology to traditional questions about moral philosophy. Of all the intellectual fashions of the late twentieth century, it has the best claim to provide an account of human nature in the scientist’s sense of ‘nature’, for it is undeniable that every aspect of humanity has been subjected to natural selection over blind variation literally since time immemorial. If any one thing has shaped us it is evolution, and if any piece of science is going to shed light on ethical issues, sociobiology - the application of Darwinian theory to human affairs - will. Therefore, my aim will be to identify the minimal conditions under which evolutionary biology *might* be able to tell us something about traditional issues in moral philosophy. If the rather strong assumptions evolutionary biology requires to shed light on these issues fall to obtain, then - as our best guess about human nature - biology will have no bearing on moral philosophy. This, in fact, is my strong suspicion. Nevertheless, I herewith attempt to put together the best-case scenario for the ethical significance of evolutionary biology

1. The possible projects

There are several sorts of insights evolutionary biology might be supposed to offer about human

² E.O. Wilson, *On Human Nature* (Cambridge: Harvard University Press, 1978).

nature and its relation to morality. One among them is uncontroversial and beyond the scope of moral philosophy. Like any scientific theory, evolutionary biology may well provide factual information that, together with independent normative principles, helps us make ethical decisions. It may uncover hitherto unnoticed means we can employ in meeting ethically established ends. It may even identify subsidiary goals that we need to meet in order to attain other intrinsic goals. For example, there are plain facts (about, for example, ecology, genetic diversity, and the importance to us of preserving threatened species) which biology reveals and which can be combined with moral standards into hypothetical imperatives governing human action.

More controversially, evolutionary biology may reveal constraints and limitations on human behavior that our ethical prescriptions will have to take account of. If 'ought' implies 'can', the contrapositive will be valid too: 'can't' should imply 'need not'. Like other scientific theories, evolutionary biology may help fill in the list of what we (nomologically) cannot do. However, for a theory of human nature to have ramifications for moral philosophy itself, it will have to do more than any of these things.

The most impressive accomplishment for a theory of human nature would be the derivation of particular moral principles, like the categorical imperative or the principle of utility, from biological facts about human beings. Slightly less impressive would be to derive from such facts our status as moral *agents* and subjects, or to establish on the strength of our biology the *intrinsic value* of human life. A derivation of agency or intrinsic value is equivalent to deriving the generic conclusion that there is some normative principle or other governing our actions. Such a derivation would be less impressive because it would leave open the question of which moral principles about agents or objects of intrinsic value were the right ones. Still less impressive but significant in its own right would be the derivation of some important component or condition or instance of morally praiseworthy conduct - like cooperation, altruism, or other-regarding behavior - as generally obligatory. To be significantly interesting the derivation need not be deductive, but it cannot be question-begging: it cannot begin from assumptions with substantial normative content. Otherwise, it will be open to the charge that these assumptions are doing all the real work, and that the biological theory makes no distinctive contribution to the derivation.

The possibility of this project, of deriving agency and/or value (or, equivalently, deriving the
Do not quote without approval of author alexrose@duke.edu

existence of some moral principle or other), rests on two preconditions. The first is that we can derive ‘ought’ from ‘is’: that there is some purely factual, empirical, contingent, strictly biological property of organisms, which could underwrite, explain, or justify their status as agents or loci of intrinsic value. The second is that this property is *common and peculiar* to *all Homo sapiens*, so that it will count as constituting our nature.

That the first of these two preconditions for deriving morality from human nature cannot be realized seems to me to be at least as widely held a view as any other claim in moral philosophy or metaethics. Accordingly, I will not offer new arguments to supplement the observations of Hume and Moore. I recognize, however, that the more sophisticated sociobiologists are perfectly aware of these strictures on moral justification. Among sociobiologists, those who nevertheless go on to attempt to derive some normative claims from biological findings do Moore and Hume the courtesy of noting and rejecting their arguments.³

But even if we grant the sociobiologist’s claim that the derivation of ‘ought’ from ‘is’ has not yet been totally excluded, there remains a second precondition required by the project of deriving morality from human nature. And the failure of this condition is something on which all evolutionary biologists should be in agreement.

Humans are supposed to be moral agents. This is what distinguishes us from moral subjects, like animals, and from morally neutral objects. Now, for some biological property of human beings to ground our status as the unique set of moral agents (in our biosphere at least), that property will have to be as widely distributed among human beings as the moral property it grounds, and it will have to be peculiar to humans as well. For if it is not restricted to humans, there will be other subjects with equal claim to the standing of moral agents. The trouble is that if modern evolutionary biology teaches anything, it shows that there are no such properties common and peculiar to each member of a species. If there were, taxonomy would be a much easier subject. And since there are none, what evolutionary

³ *ibid.*, ch.1; R. Alexander, *Darwinism and Human Affairs* (Seattle: University of Washington Press, 1979).

biology in fact shows is that there is no such thing as *the* unique human nature, any more than there is beaver nature or dodo-bird nature or E. coli nature.

Population genetics and molecular biology have shown that, up and down the entire range of living things, there are no interesting *essential* properties - no properties which will explain a range of behavior in the way that, say, molecular structure explains most of what a chemical compound does. It is not that modern biology has yet to find such essential properties, which give the nature of a species. Rather, evolutionary and genetic theory *requires* that biological species have no such common and peculiar essential properties.

Gradual evolution by natural selection requires vast amounts of *variation* within and between species. This variation is provided by mutation, genetic drift, immigration, emigration, and most of all by genetic recombination in the sexual reproduction of offspring. The result is that there are no *essential* (suites of) phenotypes. Neither the typical nor average nor mean nor median values of the heritable phenotypes which face selection are their *natural, essential* values. They do not constitute the normal traits of members of a species, from which differences and divergences might count as deviations, disturbances, defects, or abnormalities of course there are biological properties common to every member of a species. For example, all *Homo sapiens* engage in respiration. But then so does every other organism we know about. Similarly, there are some biologically-based properties peculiar to individual humans - self-consciousness, speech, a certain level of intelligence, opposable thumbs, absence of body fur, etc. But these properties are plainly not distributed universally among humans, nor would the lack of any one of them be enough to deprive someone of membership in our moral community who was otherwise endowed with it. There is no human nature in the sense in which 'natures' are identified in modern science.

It might well be supposed that there is some complex combination of properties- say,

self-consciousness *cum* opposable thumbs *cum* a disjunction of blood-types -that is sufficient for moral agency. But the project of grounding agency in (and only in) human nature requires that this complex combination of properties be necessary for agency as well as sufficient for it, and that it be universal among *Homo sapiens*. For consider, how could a property *restricted* in its instantiation only to some members of a class provide the basis for a property *common* to all members of the class: how can we derive ‘All As are Cs’ from ‘Some As are Bs’ and ‘All Bs are Cs’? Doubtless, a philosopher can solve this problem by cooking up some gruesome gerrymandered relational property. For example, one could define property C as the property of being a member of a class some of whose members have property B. Then the derivation is trivial. But, clearly, being a moral agent is not a relational property - not at any rate, if it derives solely from the nature of the individual human. And this makes the logical problem a grave one for those who seek to derive agency from human nature.

Deriving a particular moral principle, or even the generic status of moral agency, from human nature alone - at least as evolutionary biology understands it - is not feasible project, even if we could derive ‘ought’ from ‘is’.

A third potential project for the biological account of human nature is that of *explanation*: telling a plausible story about how a particular moral principle or ‘morality’ in general, or some important precondition or component of it, emerged in the evolution of *Homo sapiens*.

The *qualification* ‘plausible’ cannot be emphasized too strongly here. The most we can expect of any evolutionary account of chronology is plausibility: that the narrative will be consistent with evolutionary theory and with such slim data as may be available. The reason is that the problem of explaining the emergence of morality is similar to (but even more difficult than) that faced by, say, the task of explaining the disappearance of the dinosaur. There is a saying in paleontology: “The fossil record shows at most that evolution occurred elsewhere.” In the case of explanation the evolution of behavior, there are no bones, no “hard parts” left to help us choose among competing explanations. The most we can hope for is plausibility.

This raises the question of how much a merely plausible story is worth, what it is good for, and why we should want it for more than its entertainment value. The question is particularly pressing in moral

Do not quote without approval of author alexrose@duke.edu

philosophy and metaethics. For it is not clear that even a well-confirmed explanation for the emergence of aspects of morality from human nature has any relevance to the concerns of philosophers. It would be a genetic fallacy to infer that a particular normative conclusion was right, justified, or well grounded—or, for that matter, that it was wrong, unjustified, or groundless - from a purely causal account of its origins.

If, however, we could parlay the explanation for the emergence of aspects of morality from human nature into an argument about why it is rational to be moral, then for all its evidential weakness, the causal story would turn out to have some interest. It would address a traditional question in moral philosophy: why should I be moral? There may be some reason to think such a strategy will work. For natural selection is an optimizing force for individuals,⁴ and so is self-interest. Explanations in evolutionary biology proceed by rationalizing an innovation as advantageous for an organism's survival. Egoistic justification does something quite similar. Like evolutionary explanation, it rationalizes actions as means to ends.

This, I think, is the only interesting project in moral philosophy or metaethics for a biological approach to human nature. In what follows, I sketch the outlines of such a project. I should note two things about my sketch. First, little that follows is original. Mostly, I have plucked insights from a bubbling cauldron of sociobiological and evolutionary theorizing. Second, I am by no means optimistic that this project of rationally justifying morality can succeed, even in part. My aim is to identify the strictures it will have to satisfy if it stands a chance of succeeding.

2. Natural selection, blind variation, fitness maximization

If the theory of natural selection is right, then the overriding fact about us is that we are all approximate fitness-maximizers. Of course, this is not a special feature of people. As is explained below, natural selection cannot operate to optimize the properties of groups, as opposed to individuals. See section 11. Indeed, it is the most widely distributed property of biological interest that there is. Every organism in every reproducing species is an approximate fitness-maximizer, for natural selection selects for fitness-maximization *uber haupt*. All the phenotypes that have been selected for in the course of evolution have this in common. And if the theory is correct, then over time, given constant

Do not quote without approval of author alexrose@duke.edu

environments, successive generations of organisms are better approximations to fitness-maximization than their predecessors. This is what adaptation consists in.

What exactly is fitness-maximization? This is a vexed question in the philosophy of biology. For present purposes it will suffice to adopt the following definitions: x is fitter than y if, over the long run, x leaves more fertile offspring than y . Thus, an organism maximizes its fitness if it leaves the largest number of fertile offspring it can over the long run. It will be convenient if we define 'offspring' in a special way: an offspring will count as one complete set of an organism's genes. Therefore, the result of asexual reproduction is one offspring, but the result of sexual reproduction is half an offspring, since each child bears only one-half the genes of each of its parents. Note that, by this means of reckoning offspring, if a childless woman's brother has one child, the woman has a quarter of an offspring. Thus five fertile nieces and nephews make for greater fitness than one child: $5/4 > 1$. This means that nature, in its relentless search for fitness-maximizing organisms, sometimes selects for fewer children and more offspring.

Nature selection has made us *approximate* fitness-maximizers, not perfect ones. There are several reasons for this. To begin with, nature selects for fitness-maximization only indirectly, by seeking adaptive phenotypes: among giraffes it selects for long necks, among cheetahs for great foot speed, among chameleons for mimicry, and among eagles for eyesight. But each of these is selected because it makes for the survival and the well-being of the organisms endowed with it. And survival, along with well-being, are in turn necessary conditions for reproductive success. Mere survival is not enough; an organism must be healthy enough to reproduce and ensure the survival of its offspring. But the point is that except where selection operates directly on the organs of reproduction, birth, feeding, and protection, every other piece of an organism's equipment is selected for direct effect on survival and well-being, and through them for indirect effects on fitness. This means that much of what nature selects may not look like it bears on reproduction and fitness.

In its culling of these properties that bear indirectly on fitness, natural selection puts a premium on quick and dirty solutions to the problem of fitness-maximization. It prefers these cheap, imperfect solutions to slow but sweet ones that may do the job better but take a long time to emerge. Nature
Do not quote without approval of author alexrose@duke.edu

recognizes Keynes's maxim that in the long run we are all dead, and it acts on this maxim before it's too late. Thus, all organisms are at best approximate, jury-rigged, only intermittent fitness-maximizers. As genetic recombination and the other sources of phenotypic novelty turn up variations, the best among them out-reproduce the others. But the best may not be very good on any absolute scale. It need only be good enough to survive and outlive the other variants among which it emerges.

Selection operates on what variation provides. It has no power to call forth solutions to problems of adaptation, only to pick and choose among those that recombination and mutation may offer. Here is a nice example (with thanks to Daniel Dennett). Fish need to be able to recognize predators. But fish do not have very sophisticated predator recognition capacities; they have not evolved the sort of cognitive capacities for discriminating other fish, let alone telling friend from foe. Yet in the presence of predators they invariably startle, turn, and flee. Of course they also respond this way to all fish, not just the predatory ones- Indeed, present a fish with any bilaterally symmetrical stimulus and it will emit this flight response. The reason is that selection has resulted in the emergence of a relatively simple solution to the predator detection problem: bilateral symmetry detection. This is not a very discriminating capacity, but at least it is within the cognitive powers of a fish, and it works well enough at predator detecting. Its defects are obvious - the fish wastes energy fleeing non-predators. But in its environment the cost of this imperfection is low enough, and without it fish would not have lasted long enough to give rise to those species cognitively powerful enough to do the job of predator recognition any better.

There is a related reason why natural selection leads to the evolution only of approximate fitness-maximizers. Environments change, and organisms must survive in an environment that manifests wide extremes, and they must survive when one environment is displaced by another environment. Such conditions put a premium on being a jack of many trades instead of a master of one. An environment of great uniformity lasting over epochs of geological length provides selection with the opportunity to winnow successive variations to remarkable degrees of perfection. Consider the human eye, which is the result of a series of adaptations to a solar spectrum that has remained constant for almost the whole of evolutionary history. Such fine-tuning, however, gives hostages to fortune. For when an environment changes, there is too little variation in the received phenotype for selection to operate on. A phenotype

Do not quote without approval of author alexrose@duke.edu

that maximizes fitness perfectly in one environment is so closely adapted to it that it may not retain enough variation to survive in any other environment.

Since we are the products of selection over changing environments, we are only approximate fitness-maximizers. Nature has produced us by selecting from what was immediately available for shaping to insure short-term survival. Doubtless, in its impatience nature has nipped in the bud potential improvements in our own species and in its predecessors. For the moment the only moral of this part of the story, for moral philosophy, is this: merely showing that altruism or other well-established patterns of morally praiseworthy action are strictly incompatible with monomaniacal, perfect, complete fitness-maximization is a poor argument for the claim that human behavior has become exempt from evolutionary selection. For we are not perfect fitness-maximizers. Natural selection has shaped us for *only approximate* fitness-maximization in the environments in which *Homo sapiens* has evolved. Approximate fitness maximization leaves a great deal of room for non-adaptive altruism and other selfless actions.

3. Parameters, strategies, and the maximization of fitness

Now, among approximate fitness-maximizers, what sort of social behavior should evolve? This is a problem that arises with the advent of selection for living in family groups, which of course obtained long before *Homo sapiens* emerged. Until this point fitness maximization is, in the game theorist's terms, 'parametric', not 'strategic'. Which behavior is maximizing depends only on the environment, which provides parameters fixed *independently* of which behavior the organism is going to emit. But when organisms interact, which behavior one emits may be a function of what the other is going to do. So which behavior is fitness-maximizing will depend on how other organisms behave. This means that the optimal behavior is one that reflects a *strategy*, which takes account of the prospective behavior of other organisms. When social interaction emerges, fitness-maximization becomes a strategic problem.

This does not mean that, once groups emerge, organisms begin to calculate and select strategies based on recognition of the strategies of other organisms. It means something much less implausible. It means that those behaviors will emerge as fitter which, as a matter of fact, are coordinated with one another in the way they would have been, had they been the result of reflection and deliberation. This is *Do not quote without approval of author alexrose@duke.edu*

because there is enough time for fitness differences between the rarer *but fortuitously* coordinated behavioral phenotypes and more common uncoordinated ones to pile up and select the coordinated ones.

Coordinated behaviors are sometimes cooperative ones; they are other-regarding, involving putting oneself at the mercy (or, at least, at the advantage) of another. Thus, they constitute a significant component of morality. The emergence of coordinated behaviors makes one sort of scenario for the emergence of morality tempting. This is the *group selection* scenario, according to which nature selected societies and groups because their institutions, including their moral rules, are more adaptive for the group as a whole. On this model, selection proceeds at the level of the individual (for individual fitness-maximization) and at the level of the group (for group survival and growth). The idea that evolution might lead to the emergence of morality by selecting for groups that manifest moral rules and against groups that retain a state of nature is, on the face of it, more attractive than trying to find a story of how morality might have emerged at the level of the individual. For the fitness-maximizing individual is concerned only with maximizing its offspring; it is ready to sacrifice others to this end. The fact that there are so many immoral and amoral people around makes implausible the notion that morality emerged among *Homo sapiens* the way opposable thumbs did -as an individual response to the selection of individual organisms. But the emergence of morality as a group institution is at least compatible with the observed degree of moral imperfection among individuals. Selection at the level of the group does not require uniformity among- the individuals who compose it; no championship team has ever had the best players in the league at every (or even any) position (cf. New York Mets, 1969).

So, one way to reconcile other-regarding cooperative behavior with monomaniacal evolutionary egoism is to locate selection for cooperative institutions at the level of the group and selection for individual fitness-maximization at the level of the individual. If the forces selecting for the adaptation of groups are independent of those selecting for the adaptation of individuals, then those groups within which cooperation, promise-keeping, property, fidelity, etc., emerged, for whatever reason, might do better, last longer, or have larger, healthier populations than those groups which lacked such virtuous institutions. Thus, morality is explained as an evolved holistic social constraint on individual selfishness.

Do not quote without approval of author alexrose@duke.edu

This is a nice idea, but one which evolutionary biology must exclude. For no matter how much better off a society with ethical institutions might be than one without them, such a society's seriously unstable, and in the evolutionary long haul must fall victim to its own niceness. The reason is that selection at the level of the group and the level of the individual are never independent enough to allow for the long-term persistence of a moral majority and an immoral minority. In fact, they aren't independent at all. The latter will eventually swamp the former.

Consider a society of perfectly cooperative altruistic organisms, genetically programmed never to lie, cheat, steal, rape, or kill, but in which provisions for detection and elimination of organisms who do not behave in this manner are highly imperfect (as in our own society). Since everyone is perfectly cooperative, the society needs no such provision. Now suppose that a genetically programmed scoundrel emerges within this society (never mind how - it might be through mutation recombination, immigration, etc.). By lying, cheating, stealing, raping, and otherwise free-riding whenever possible (recall the detection and enforcement mechanisms are imperfect), the scoundrel does far better than anyone else, both in terms of well-being, and in terms of eventual fitness-maximization. He leaves more offspring than anyone else. If his anti-social proclivities are hereditary, then in the long run his offspring will come to predominate in the society. Eventually, 100 percent of its membership will be composed of scoundrels and its character as a cooperative group will long since have disappeared.

Evolutionary game theorists have provided a useful jargon to describe this scenario: a group with a morally desirable other-regarding strategy is not "evolutionarily stable": left alone, it will persist, but it can be "invaded" even by a small number of egoists - who will eventually overwhelm it and convert the society into one bereft of other-regarding patterns of interaction. By contrast, a society composed wholly of fitness-maximizing egoists is an "evolutionarily stable" one: a group of such egoists cannot be successfully invaded by some other, potentially nicer pattern of behavior. Its members will all, one after another, play the nice guys for suckers and out breed and ultimately extinguish them.

The trouble, then, with group-selectionist explanations of the emergence of morality is that a group of other-regarders might do better than a group of selfish egoists, but it is vulnerable to invasion by one such egoist, an invasion which evolutionary theory tells us must always eventually occur - since nature is *Do not quote without approval of author alexrose@duke.edu*

always culling for improvements in individual fitness-maximization. Whether from within or without, scoundrels will eventually emerge to put an end to other-regarding groups by converting them into societies of fitness maximizers.

If morality is to emerge from the nature of organisms as approximate fitness-maximizers, it will have to happen at the level of individual selection. And it will have to be selection for optimizing behavior in the context of “strategic” interaction, where the optimum behavior of each organism depends on the behavior of other organisms. The trouble is that game theorists have increasingly come to suspect that there is no optimal strategy under these circumstances. If this is right, then there is none for evolution to choose, and no way for moral institutions to evolve from the strategic interactions of fitness maximizers.

The problem of an optimal strategy for nature to select is easily illustrated in the children’s game of Rock, Paper and Scissors. In this game, kids pick one of the three choices. Rock breaks scissors and so beats it, scissors cut paper and so beat it, but paper covers rock and so beats it. Whether your choice wins depends on what the other kid picked, and no choice is better than any other. In an evolutionary situation like this, no strategy ever comes to predominate. Of course, if you know what your competitor will pick, you can always win. But what the other kid picks is going to depend on what he thinks you will pick. So, you have to know what he thinks you will pick in order to pick the best strategy, and so on backwards *ad infinitum*. There is no end to the calculation problem, and therefore no optimal strategy in the rock-paper-scissors game. Game theorists have labeled the problem of this sort of game with no finite solution - no best strategy for any player - "the problem of common knowledge." In principle, the problem of having to infinitely iterate calculations about what other players will do bedevils most strategic games.

While the problem of common knowledge cannot affect organisms which are incapable of making calculations about the strategies of others, it can affect the evolution of fitness-maximizing strategies. As nature selects the best among competing strategies for fitness-maximization, it must eventually face contexts in which the best strategy for an organism to play depends on what other strategies are available to be played by other organisms. If the game theorist can prove that, in the long run, there is no single best strategy - even with rational calculation on the part of the players - then we can expect

Do not quote without approval of author alexrose@duke.edu

natural selection to do no more than produce a motley of equally good or bad strategies that compete with one another, at best gaining temporary ascendancy in a random sequence. In other words, natural selection will produce nothing but noise, disorder, no real pattern in the behavior of fitness-maximizers who face strategic competition as opposed to parametric optimization problems.

It seems safe to assume that *Homo sapiens* has not in fact suffered this fate. For the most part, our interactions do show a pattern, and an other-regarding, cooperative one at that. Morality is the rule and not the exception (and not just one among a series of cyclically succeeding patterns of behavior). It must follow, therefore, that evolution has not led us (or our evolutionary forbearers) down the cul-de-sac of the problem of common knowledge. But, if game theorists are right, almost the only way evolution could have avoided this sort of chaos is for other-regarding principles of conduct to have emerged in parametric contexts, and then to be evolutionarily stable, un-invadable when these contexts became strategic.

4. Kin-selection and uncertainty

For a fitness-maximizing organism, interactions with offspring are close to being parametric. For, almost no matter what children and kin do to you, if you act in their interests, the result will increase your fitness. The fitness-maximizing strategy for an organism is therefore to act so as to maximize the fitness of its offspring. Thus, in selecting for fitness-maximization, nature will encourage organisms whose genetically encoded dispositions include sharing and cooperating, and even unreciprocated altruism towards kin - children, siblings, even parents. For these strategies are likely to increase one's offspring, no matter how they respond to you. This sort of kin-altruism is evolutionarily stable and un-invadable. A short-sighted, selfish organism, who behaves as though its own survival or that of its children counted for its fitness, would end up with fewer offspring over the long haul. For sometimes it would look out for number one (or number one's kids) when sacrificing itself or a child would result in the survival of a larger number of offspring (recall, that under sexual reproduction, a child is only half an offspring). In selecting for fitness, nature will select for 'inclusive fitness' and 'kin selection' will emerge. Kin selection is something we can count on emerging long *before* *Homo sapiens* appears. It becomes an adaptive strategy as soon as the number of genetic offspring begins to exceed the number of children.

Do not quote without approval of author alexrose@duke.edu

(Recall, three nephews carry more of an individual's genes than one child.)

When *Homo sapiens* emerges, therefore, we are already beyond Hobbes's state of nature. Cooperation, altruism, and other-regarding behavior generally is already established inside both the nuclear and the extended family. Indeed, it is likely to already have been established a bit beyond this. Consider that individuals do not wear name tags or carry their genealogy on their sleeves for others to examine before deciding whether an interaction will be parametric or strategic. There are, of course, clear signs of kinship that even animals with limited recognition powers can use: odor, proximity to a nest or region. And there are clear signs of xenonimity-- strangeness. But there is always a large area of uncertainty in between, a range of interactions in which two organisms just can't tell with any more than a certain moderate level of probability whether they are kin or not. This will be more true for males and their putative offspring than for females. Given the nature of procreation and gestation in many mammalian species--and especially in *Homo sapiens*-- the male can never be as certain as the female that the young in his family are his offspring--i.e., that they share some of his genes. Unless the female is under constant and perfect surveillance during the critical period, the question of whose sperm fertilized her ovum must always be a matter of some doubt. Beyond the relation between mother and child, the degree of consanguinity between any two organisms is always a matter of probabilities, and doubts about kinship are easier to raise than to allay.

Under conditions of uncertainty about kinship, what is the optimal strategy for a fitness maximizer? Game theory tells us that the rational thing to do is to apportion the degree of one's other-regarding behavior to the strength of the evidence of consanguinity. In the long run, as natural selection operates, it must favor this strategy as well. Even in cases where the available positive evidence of consanguinity (subtle similarities of smell, coat, color, shape of beak, pitch of mating call, etc.) is difficult to detect, one can expect nature to select for cooperation and other-regarding behavior between kin, provided only that it has enough time to fine-tune the detection mechanisms. Considering the job it has done in optimizing the eye for vision within the time constraint of four million years, it may seem reasonable to suppose it can fine-tune kin-selection strategies as well. And if everyone turns out probably to be closely enough related to everyone else, then natural selection might be expected by itself to produce

Do not quote without approval of author alexrose@duke.edu

other-regarding behavior up to levels of frequency that match the probability of universal consanguinity. Here we have the emergence of morality, or at least a crucial aspect of it, without having to solve the problems common knowledge makes for strategic games. However, the amount of other-regarding behavior that might in fact be fitness-maximizing just because of the fact that we are all each other's seventh cousin hardly seems sufficient to explain the emergence and persistence of moral conduct. The problem with this neat explanation is that we have no independent idea of whether the payoffs (in more offspring) for being other-regarding are really great enough when the probability of being related falls to the level that obtains between you and me. And there doesn't seem to be any easy way to find out. In short, our explanation isn't robust enough. It rests on a certain variable taking on a very limited range of values, one within which we have no reason to think it falls. We need a better explanation for the emergence and persistence of other-regarding behavior than kin selection and the uncertainty of relatedness can give us. It's all right to start with kin selection, but we need an explanation that carries other-regarding conduct into the realm of strategic interactions among fitness maximizers *unlikely* to be kin.

To do this, we need to help ourselves to another brace of healthy assumptions about morality and game theory. First, let us accept without argument that the institutions of morality are public goods: they cannot be provided to one consumer without being provided to all others, so that any one consumer has an incentive to understate the value of the good to him and so decline to pay its full value provided he is confident that others will pay enough to provide it. Certainly, the institution of generalized cooperation is like this. No one can count on it unless everyone can, and we all have an incentive to understate its value to us whenever we are asked to pay our fair share to maintain it. Moreover, fitness-maximizers have an incentive to cheat, to decline to cooperate, if they can get away with it undetected or unpunished. But if everyone knows this, and everyone knows that everyone knows this, etc., then the institution of cooperation will break down because of our common knowledge. The public good is lost, and every one is worse off. The prisoner's dilemma graphically illustrates this problem of the provision of public goods. Individual rational agents have an incentive to be free-riders, to decline to cooperate. The result is a non-optimal equilibrium in which no cooperation is visible. The natural selection version

Do not quote without approval of author alexrose@duke.edu

of this collapse from the fortuitous provision of public goods to a non-optimal equilibrium takes time, as individual defectors emerge through recombination or mutation and out-reproduce cooperators.

The second assumption we need is that most of our morally relevant interactions are moves in an indefinitely long sequence of prisoner's dilemma games. This seems a not unreasonable assumption: honoring moral obligations is not a one-shot, all-or-nothing affair. It is a matter of repeated interactions largely among the same individuals. Interactions with strangers are by definition less frequent than with people we have interacted with and will interact with in the future. Now, one important fruit of the joint research of game theorists and evolutionary biologists has been the conclusion that, even among strangers, being a free-rider (always declining to cooperate, always taking advantage) is not the fitness-maximizing strategy in an iterated prisoner's dilemma. Rather, the best strategy is what is known as "tit-for-tat": that is, for optimal results one should cooperate on the initial occasion for interaction, and on each subsequent occasion do what the other player did on the last round. This strategy will maximize fitness even when everyone knows that everyone else is employing this strategy. For even on the assumption that there is complete common knowledge of what strategies will be chosen, tit-for-tat remains the best strategy. Once in place, it assures cooperation even among unrelated fitness-maximizers. It circumvents the common-knowledge problem.

5. Ethics - quick and dirty

There is one rather serious obstacle to natural selection's helping itself to this strategy: the problem of getting it into place. For tit-for-tat cannot invade and overwhelm the strategy of narrow selfishness that is required by strict fitness-maximization. In a group of organisms that never cooperate, anyone playing tit-for-tat will be taken advantage of at least once by every other player. This advantage is enough to prevent tit-for-tat players eventually swamping selfishness. In fact, it may be enough of an advantage for tit-for-tat to be driven to extinction by the strategy of selfishness every time it appears as a strategy for interaction.

This is where nature's preference for the quick and dirty, *approximate* solution to the problem of selecting for fitness-maximization comes in. Our approximate fitness-maximizers' optimum strategy involves other-regarding behavior with kin, and selfishness with others. How will fitness be maximized
Do not quote without approval of author alexrose@duke.edu

in the borderline area where kinship and its absence are difficult or impossible to determine? When the only choice for an organism is to cooperate or decline to do so, how does it behave? By flipping a coin weighted to reflect the evidence for kinship, and doing as the coin indicates? A few pages back I derided this suggestion, though we cannot put it past nature to have evolved a device within us that has this effect. On the other hand, nature will prefer quick and dirty solutions to mathematically elegant ones, provided they are cheap to build, early to emerge, and do the job under a variety of circumstances, etc. If tit-for-tat is almost as good a strategy for fitness-maximization in cases of uncertainty as employing the probability calculus and far easier for nature to implement, then on initial encounters under uncertainty about kinship, individuals playing this strategy will cooperate. But this means they will cooperate thereafter as well. Thus, interactions at the borderline come to have the character of interactions within the family; parties to any and every interactive situation will generally cooperate.

Now suppose that among organisms genetically programmed to be other-regarding within the family and to play tit-for-tat at the borderlines, one or more individuals emerge with a new variation: their genome is programmed to encourage tit-for-tat always and everywhere, or at least whenever interacting with strangers. Interaction with selfish strangers will be costly to such organisms and should lead to their extinction. But suppose such interaction is rare. Furthermore, suppose (as seems reasonable) that the strategy of always playing tit-for-tat is otherwise an adaptive one, with advantages over other more complex strategies, especially for organisms lacking complex cognitive and calculational powers. For the cost of maintaining and using a storage system for kin and non-kin may be greater than the cost of being taken for a sucker in just the first round of an indefinitely iterated interaction. This will likely be true when the chances of meeting a stranger are extremely low, as they will be in the earlier stages of the evolution of mammalian species living in family groups. It is, in general, easy to imagine scenarios that make tit-for-tat the best overall strategy under most circumstances in a given environment. But this means that natural selection for approximate fitness maximization among individuals has led to the emergence of cooperative, other-regarding strategies. It has solved the problem of providing public goods to individual organisms geared always and only to look out for themselves and their kin. If ethical

Do not quote without approval of author alexrose@duke.edu

institutions are, after all, public goods, then we have explained how they might emerge among approximate fitness-maximizers.

Of course, this entire story applies to us only to the extent that we are approximate fitness-maximizers. This is not hard to show. In fact, if anything, it's too easy to show. For the story does not include any indication of how good an approximation to perfect fitness-maximization is required for the emergence of other-regarding strategies. Even if it did, we have no idea of whether *Homo sapiens* is in fact a good enough fitness-maximizer for this scenario actually to obtain. For these reasons, the claim that we are in fact approximate fitness maximizers will have vanishingly small empirical content. But then empirical content was never the strong point of any evolutionary theory, and is of little interest in moral philosophy anyway.

That we are fitness-maximizers to some degree of approximation goes without saying. After all, the only alternative to being an approximate fitness-maximizer is being extinct. And how did nature shape us for fitness-maximization? What phenotypical properties of *Homo sapiens* did it shape in this direction? Well, the quickest and dirtiest way of making us approach fitness-maximization is to make *us approximate utility-maximizers*, to shape us into systems organized to maximize our well-being, by linking well-being to the avoidance of discomfort, pain, and distress, and the attainment of comfort, pleasure, and feelings of security. The reason is obvious: an organism's reproductive potential is, *ceteris paribus*, a function of its well-being. So, in order to select for fitness-maximization, nature will select for organisms that by-and-large maximize their well-being. The by-and-large clause reflects the fact that there are certain departures from utility-maximization that nature will select for too. For example, it will select for organisms that sacrifice their own well-being to offspring, especially after they have passed the a(re of optimal procreation. Or, equivalently, nature will select for preference structures that make kin-altruism pleasing to the individual. This is a quick and dirty solution to the problem of programming kin-selection, one which corresponds to the philosopher's claim that altruism is just the reflection of a perverse preference structure.

If the quick and dirty solution to the problems of designing an approximate fitness-maximizer is to design an approximate utility-maximizer, then our merely plausible explanation for the emergence of

Do not quote without approval of author alexrose@duke.edu

morality or of one important component of it may have another role to play. It may turn out to be a part of a (weak) *justification* of morality, or at least of one important component of it.

One traditional question of interest to moral philosophers is that of how to convince the rational egoist to be moral, how to show the egoist that being moral is in his interest. Nowadays this problem is often set forth as that of showing how morality could be part of a strategy that maximizes individual utility. In its own way, natural selection provides reason to suppose that morality is part of a utility-maximizing strategy, and our story provides a plausible scenario for how this might have happened.

It is clear that nature began selecting for utility-maximizers long before it began selecting for other-regarding cooperators. For one thing, maximizing well-being is a strategy to be found across the phylogenetic spectrum; it doubtless characterized our ancestors long before the rearing of offspring in nuclear or extended families and the emergence of social groups made other-regarding cooperation possible and necessary. Having laid down very early in evolution approximate utility-maximization as a quick and dirty strategy for approximating fitness-maximization, nature is unlikely ever to "rip it out" and start over. This means that when it lays down other strategies, they will at least have to be compatible with utility maximizing. It's much more likely that the new strategy will be new ways to maximize utility under new circumstances. But if cooperation and other-regarding behavior generally is nature's way of most efficiently maximizing utility, then it should be good enough for us. That is, in our own calculations and reflection on how to maximize our utilities, we should expect to come eventually to the same conclusion which it has taken nature several geological epochs to arrive at. Both rational agents and nature operate in accordance with principles of instrumental rationality; they both seek the most efficient means to their ends. Since nature's end (approximate fitness-maximization) is served by our ends (approximate utility-maximization), our means and nature's will often coincide.

6. Conclusion

It's a nice story, and it seems to have a moral for moral philosophy. I think it is absolutely the best biology can do by way of shedding light on anything worth calling 'human nature' and drawing out its implications for matters of interest to moral philosophers. But before taking any comfort in it at all, we *Do not quote without approval of author* alexrose@duke.edu

need to recall and weigh the hostages to fortune it leaves - the many special assumptions about us and about the nature of moral conduct that it requires just to get off the ground: to begin with, the idea that just because one is cooperative or other-regarding, one has attained the status of a moral agent or some important precondition to it. Then there are the claims about humankind as approximate fitness-maximizers. Even if you accept this view of 'human nature', as I do, you are committed to a level of fitness-maximization that you cannot specify beyond saying it is high enough to allow for the scenario I have tried to unfold. Then you have to find a way to draw the force or circumvent the difficulty of the problem of strategic games, in which there seem to be no stable equilibria in the behavior of fitness-maximizers, let alone equilibria that underwrite any part of morality. (And you can't call upon group selection to help solve this problem.) Then you have to buy into the theory of kin-selection and its application to conditions of uncertainty. This is one of the smaller gnats to strain on, given the independent evolutionary evidence for kin selection. But the trouble is that it will not suffice when interactions begin to transcend the family. At this point, we need to assimilate morality further to strategies of choice to be analyzed by the tools of economics and game theory. Finally, we need to be able to fudge our account enough to say that morality emerges because we are not perfect fitness-maximizers, since the best nature can do is make us approximate utility-maximizers.

But perhaps the most difficult consequence of this story to swallow is this: if nature had been able to do any better, morality might never have emerged at all.